AD-A253 343

PROGRESS REPORT

For the Period Ending July 1, 1992

VISUAL PERCEPTION OF DEPTH-FROM-OCCLUSION: A NEURAL NETWORK MODEL

Contract # N00014-90-1864

P.I - Leif H. Finkel
Department of Bioengineering and
Institute of Neurological Sciences
University of Pennsylvania
220 South 33rd Street
Philadelphia, PA 19104-6392

PROGRESS REPORT — July 1, 1992

VISUAL PERCEPTION OF DEPTH-FROM-OCCLUSION: A NEURAL NETWORK MODEL

During this period we have made continued progress in simulating intermediate-level visual processes. We have applied our object-discrimination system to real video images. The model successfully extracts depth-from-occlusion in real images, as well as in a variety of "illusory contour" stimuli. In addition, we have extended a new model of texture discrimination to the problem of determining shape-from-texture. Finally, we have extended our model of color vision to account for many of the classical effects in color contrast and color constancy.

This report covers progress in the three-month period since our last report.

Model of Depth–from–Occlusion

Results with Real Images

We have begun to test our system with real video images. A video system has been constructed consisting of a Pulnix CCD video camera and an Imaging Technology S151 image processor. We have written extensive software for a range of image processing applications on the S151. The image processor is connected to our SUN workstation network, and provides direct input to the NEXUS neural simulator. Figure 1 shows the results of the depth-from-occlusion model for a real image consisting of a pen behind a styrofoam cup. The system has discriminated the occlusion boundaries in the scene, has bound surfaces and contours so as to discriminate the two objects (pen and cup), and has accurately ordered the two objects in relative depth. Note that not all contours in the image are represented in the network output, this is because the early vision networks act to select only occluding contours. Additional work is required on this point, as well as to deal with complications such as specular reflections and shadows. We are in the process of building a much more powerful early visual system in the context of the texture discrimination model discussed below. Nonetheless, preliminary tests have shown that increasing the complexity of the image poses no problems for the system.

Early Vision and Binding

We have made a number of improvements in our basic model of object discrimination based on depth-from-occlusion. The early visual networks have been made more consistent with the properties of complex cells in striate cortex. We have also developed a new algorithm for contour binding–the process which determines which points in the image belong to the same curve. The algorithm is loosely based on the notion of phase-dependent firing as observed by Gray and Singer and others. The algorithm first binds units responding to nearby points on the same line or curve, and then binds points across discontinuities (e.g. on different sides of a triangle) based on a novel gating mechanism. While most phase-dependent models have no mechanism to assure that separate objects fire at different phases, we have developed an inhibitory network which assures that successively bound contours fire at very different phases.

Results of simulations have prompted a new view of how depth-from-occlusion operates. Our initial tendency (and that of several models that have just come out) is to search for cues to occlusion–T-junctions, line endings, concavities, etc., and then to determine the relative depth of surfaces based on these cues. There are two problems with this approach: it makes depth-from-occlusion a dedicated process (as would be all other modules such as shape-from-X), and secondly, it requires that there exist neurons that detect these cues. It is important to note that except for a few early anecdotal suggestions, no one has ever found cells in striate cortex that are primarily responsive to junctions, line crossings, or any of the other cues required by such models. (However, psychophysical experiments do suggest that such cues are distinguished preattentively).

Our new approach takes the broader view that depth-from-occlusion is but one aspect of the general problem of object discrimination, and that the critical process in defining an object is the representation and binding of surfaces. Thus, in the current version of our network simulations, we determine depth-from-occlusion purely based on surface bindings. When one surface occludes another, there is an indeterminacy in the boundary between the two surfaces–whichever surface "owns" the border (to use Nakayama's terminology) is the occluding (nearer) surface. Our networks determine which surface owns the border by carrying out the process of binding contours to surfaces. Thus, depth relationships fall out of the more general process of developing an intermediate-level representation of an object.

Current Work

We are currently working on a number of other applications, including the perception of transparency, and a comparison of the model's outputs to various psychophysical results (in particular, a careful study of what makes illusory contours more or less perceptually vivid). Finally, we are beginning to incorporate recent physiological results of Gilbert and Wiesel which suggest that receptive fields of visual neurons may be dynamically plastic, and may be able to rapidly increase in size to span occluding gaps.

Model of Shape–from–Texture

Over the last several months, we have developed an energy-based model of shape-from-texture. The architecture of the network is shown in figure 2. We have used an early vision system consisting of orientation selective and center-surround units at several spatial frequency scales, based on the work of Adelson and Bergen, Malik and Perrona, and others. However, we have modified previous models by using ON and OFF centered cells. Our major point of departure from earlier models is to consider how to use the response of these early energy detectors to determine the curvature of a textured surface.

The basic idea is shown in figure 3. The textured pattern shown can be perceived as lying on a curved cylindrical surface. As a surface curves, the appearance of a geometrical pattern upon that surface undergoes two different changes. First, the *projection* of the pattern upon the retina changes, for example, it is foreshortened as it approaches the sides of a cylinder. Secondly, the *density* of patterns on the surface is altered, for example, as you approach the sides of the cylinder the density of patterns increases. We propose that these two changes are detected by monitoring changes in the energies in orientation-specific channels. For example, as shown in Figure 3, as one approaches the sides of the cylinder, the vertical energy increases and the horizontal energy decreases.[1] The basic idea is thus that curvature can be detected as an **anti-correlation** in the change of texture energies over some spatial extent. Note that any individual texture energy, i.e., the amount of 45^o lines may increase or decrease over a scene. It is the precise nature of an increase in one texture energy exactly correlated with a decrease in another (roughly orthogonal) texture energy that accurately signals surface curvature.

Figure 4 shows the results of network simulations based upon this principle. The surface curvature of both the sphere and the cylinder are detected. We have similarly shown that the network provides accurate responses to a number of other standard 3D shapes. While the basic principle of detecting changes in the distribution of energies appears robust, we are currently trying to drastically reduce the number of required networks. A major new area of research in machine vision studies of shape-from-texture involves the use of Fourier techniques. We have just completed a mathematical analysis showing that under somewhat general conditions, our approach is identical to the Fourier approach. However, we believe that our algorithm is better suited for network implementations (in addition to being biologically-motivated). In addition, our system should work for non-periodic and random textures, whereas the the Fourier approach may be best suited to periodic micro-textures.

Current Work

We are currently extending our simulations to incorporate effects of perspective changes. In addition, we are planning to include information regarding the bounding contours of the textured surface. Psychophysical evidence suggests that humans are actually not very accurate at determining shape-from-texture, and what is needed is more a qualitative notion of 3D-surface curvature in which boundary information and surface information are combined, in much the same way as developed in our depth-from-occlusion model. We are developing psychophysical tests to compare the results of our texture model with human estimates of surface curvature.

Model of Color Vision

We have developed a model of color constancy and color induction based upon the projection from retinal cones to cortical area V4. As described in previous reports, the key to this model is the computation of cone-specific color contrast in our network model of area V4. V4 cells determine contrast, or the difference in activation between the receptive field center and surround (20^o visual field). This contrast signal (which can be positive

[1] vertical energy refers to the squared normalized output of vertically-tuned orientation units. In this case, vertical energy increases due to the increased density of vertical lines; horizontal energy decreases due to forshortening of lines along the direction of curvature.

or negative, and is thus detected by two types of contrast cells) is used to modulate the output of a separate population of V4 cells which respond only to activation in a restricted receptive field location. We have found that this cone-specific contrast mechanism can account, semi-quantitatively, for psychophysical results in the perception of both color and luminance.

Color Induction

We have developed simulations testing the behavior of our network with regard to both color induction. One set of simulations deal with the dependence of color induction upon the spatial distribution of reflectances in the scene. We have tested our network system with the same stimuli that Blackwell and Buchsbaum, and Walraven and colleagues have used to determine psychophysical responses in human subjects. The stimuli consist of small colored squares inside large colored surrounds (colors are selected from Munsell chips) with a neutral grey gap between the center and surround. The size of the color induction effect has been shown to decrease as the size of the gap between center and surround increases. Figure 5 shows that the network behaves similarly to human psychophysics, with monotonic decreases in the amount of color induction as gap size increases.

Color Constancy

With regard to color constancy, the network generates "human-like" responses to Mondrian stimuli in a classic Land-McCann experiment. In this experiment, the network is presented with a color Mondrian illuminated with a standard illuminant (corresponding to noon-time sunlight) in which the center patch of the Mondrian is, for example, blue. The illuminant is then altered until the reflected wavelengths of the center patch equal that for a green patch viewed under the standard illuminant. The Mondrian under the altered illuminant is then presented to the network. If the network exhibited perfect color constancy, it should perceive the center patch as blue. If it exhibited no color constancy whatsoever, the center patch should be perceived as green. In fact, when the stimulus is shown the network, it perceives the center patch as blue-green, thus demonstrating partial color constancy. This corresponds to human behavior, as we do not exhibit perfect constancy either.

We have recently begun to test the ability of the network to match human performance on mondrians illuminated with non-uniform illuminants. Thus far, the system performs well with linearly non-uniform illuminants.

Roles of Early vs. Intermediate Vision in Color Perception

One of our most interesting findings concerns the relative contributions of cone-specific contrast versus adaptation in color constancy. Figure 6 shows simulation results for several colored spots on a grey background, viewed under two different illuminants. These illuminants differed only in luminance, not hue or saturation. As can be seen, when the luminance of the illuminant is altered (right panel), color constancy fails rather dramatically. We are led to believe that adaptation is necessary to adjust for changes in luminance while cone-specific contrast adjusts for changes in the hue-saturation of the illuminant. Adaptation probably occurs in the retina, and may contribute to hue-saturation effects (particularly after extended viewing), but the model predicts that the immediate perception of the "color" part of color constancy is due to the cone-specific contrast determined in cortical area V4. We are currently pursuing these observations with additional simulations.

Invited Presentations

Results of our work have been presented at several conferences:

- ARVO-92, Sarasota, May 1992 (two papers)
- Selectionism and the Brain, Rockefeller University, May, 1992
- International Joint Conference on Neural Networks, Baltimore, June, 1992 (three papers)
- Computer Vision and Pattern Recognition-92, Urbana, June 1992
- We have also used the NEXUS simulator to run a computational neuroscience lab at the McDonnell Institute Summer Course on Cognitive Neuroscience at Dartmouth University. In addition to presenting our own research, we trained 72 students on the use of NEXUS and had them run three simulations: organization of topographic maps (simulating the experiments of Merzenich and his colleagues on monkey somatosensory cortex), an energy-model of texture discrimination (based on the work of Adelson, Bergen, Malik and others), and a PDP model of object recognition (in which the networks were trained to classify different species of leaves, oak, maple, beech, etc.).
- Later this summer, we will make presentations at CNS*92, the Society for Computer Simulation Annual Conference, The Whitaker Foundation Annual Meeting, and the Annual Meeting of the Optical Society of America.

Publications

1. Leif H. Finkel and Paul Sajda (1992) Object Segmentation and Binding within a Biologically-Based Neural Network Model of Depth–from–Occlusion. *Computer Vision and Pattern Recognition '92*, Urbana, Ill. pp. 688-691.

2. Paul Sajda and Leif H. Finkel (1992) Simulating Biological Vision with Hybrid Neural Networks. *Simulation* [in press].

3. Leif H. Finkel and Paul Sajda (1992) Computer Simulations of Object Discrimination by Visual Cortex. *CNS*92 Proceedings*, San Francisco [in press].

4. Leif H. Finkel and Paul Sajda (1992) Proto-Objects: An Intermediate-Level Visual Representation. *Technical Digest of the Annual Meeting of the Optical Society* [in press].

5. Susan Courtney, Gershon Buchsbaum, and Leif H. Finkel (1992) Cone Adaptation and Cortical Silent Surrounds Cooperate to Produce Color Constancy and Color Contrast. *Technical Digest of the Annual Meeting of the Optical Society* [in press].

6. Paul Sajda, Ko Sakai, and Leif H. Finkel (1992) NEXUS: A Tool for Simulating Large–Scale Hybrid Neural Networks. *Society for Computer Simulation Conference Proceeedings*, Reno, Nev. [in press].

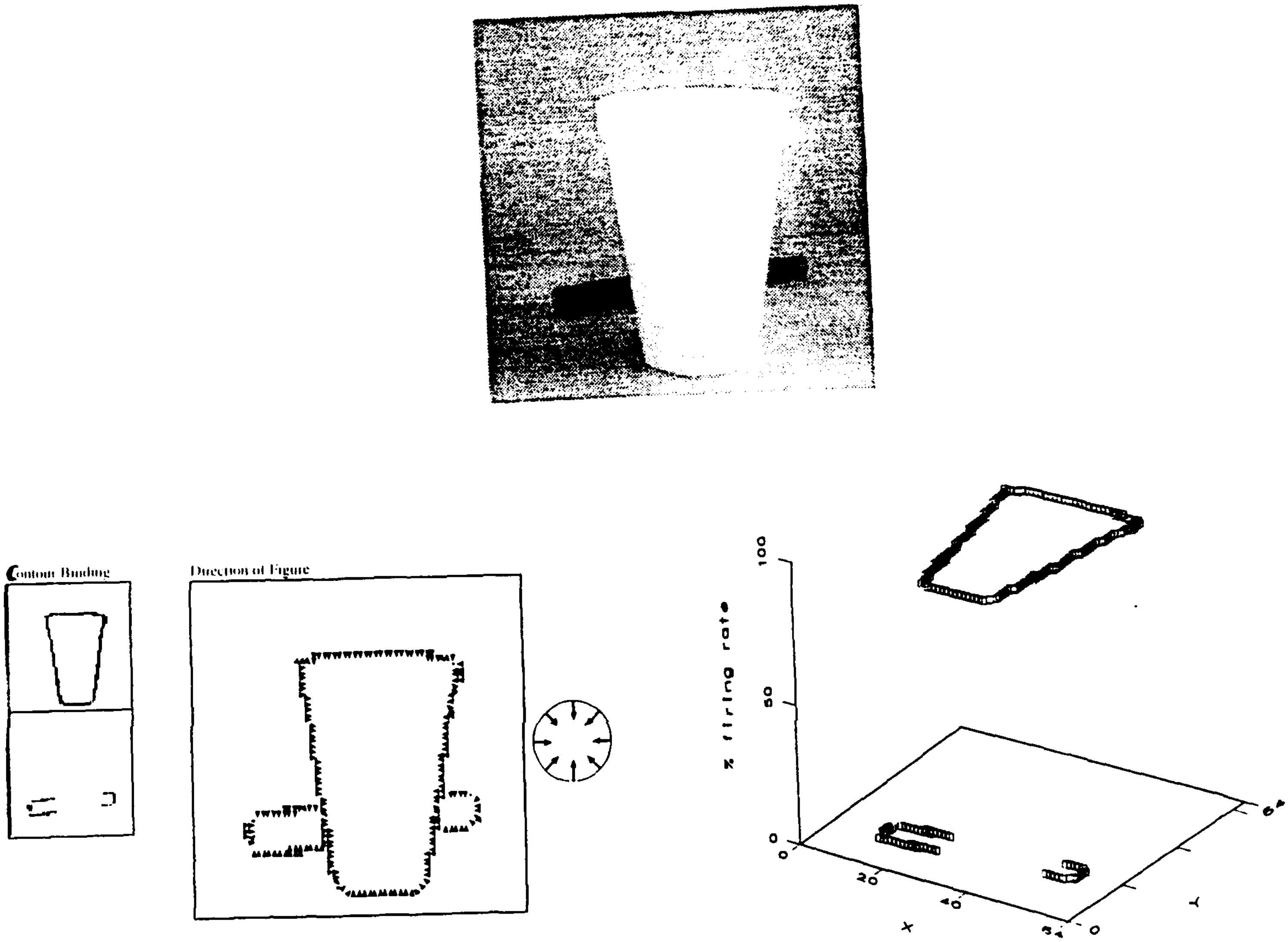

Figure 1—Example of real video image presented to network and resulting segmentation, surface binding, and depth-from-occlusion processing. **(Top)** Real image obtained with Pulnix CCD Video Camera and Imaging Technologies S151 image processor. **(Bottom Left)** Outputs of two of the 10 major networks in the model. Cup and pen have been segmented into two separate objects, as revealed by separate *contour bindings (each window represents units with a different* "phase"). Note that two ends of the pen are bound (by the same phase) despite their spatial separation due to occlusion. Direction of figure shows the direction of interior surface (arrow heads). **(Bottom Right)** Relative depth as "perceived" by the network. Plot shows firing rate of units in a network activated most strongly by nearby objects (depth is coded in a distributed fashion by units in two networks that respectively prefer nearer objects and more distant objects). Network has discriminated correct relative depth of cup and pen.

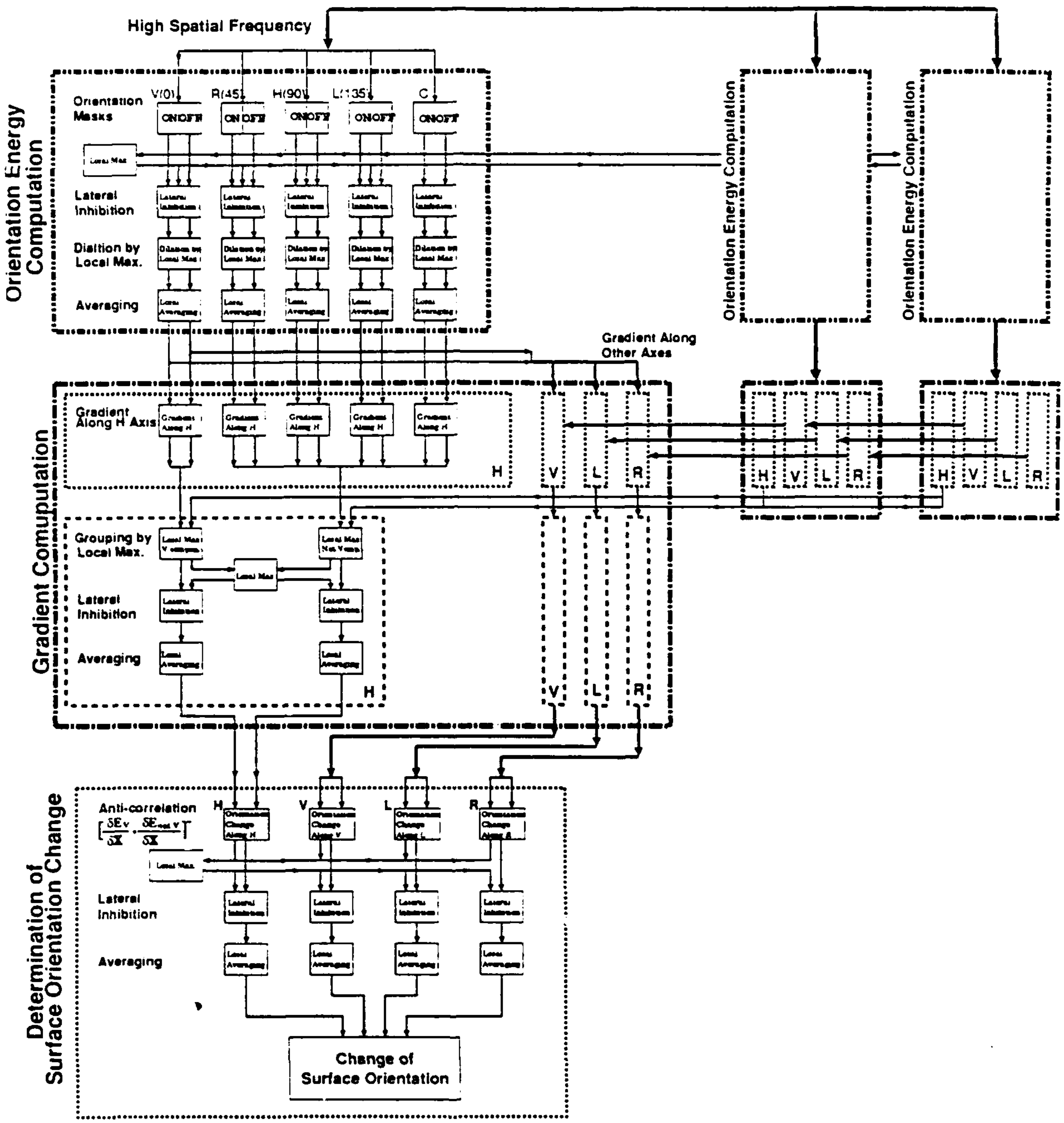

Figure 2—Schematic of shape-from-texture network. Image is sampled by orientation and circularly-symmetric energy units at three spatial frequencies. Several stages of Malik-type lateral inhibition are followed by computation of gradient of response along different directions. Surface curvature is signalled by anticorrelation of energy gradients.

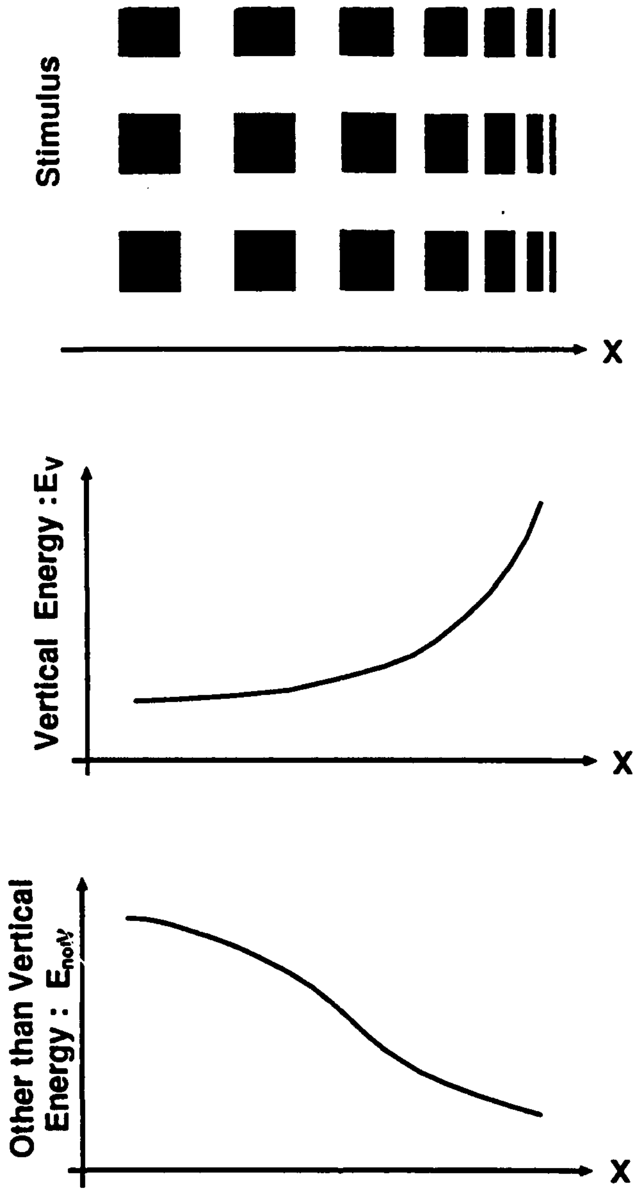

Figure 3—Basic Principle underlying Shape-from-Texture Model. Textured pattern at top can be perceived as lying on a curved cylindrical surface. Bottom two panels show the response of energy-type units responsive to vertical and non-vertical (i.e. sum of horizontal and oblique) orientations. Curvature is characterized by an anticorrelation of the change in energy for the vertical versus non-vertical units. Note that textures can change arbitrarily over a surface, but this correlated change in different components is a reliable signal of surface curvature.

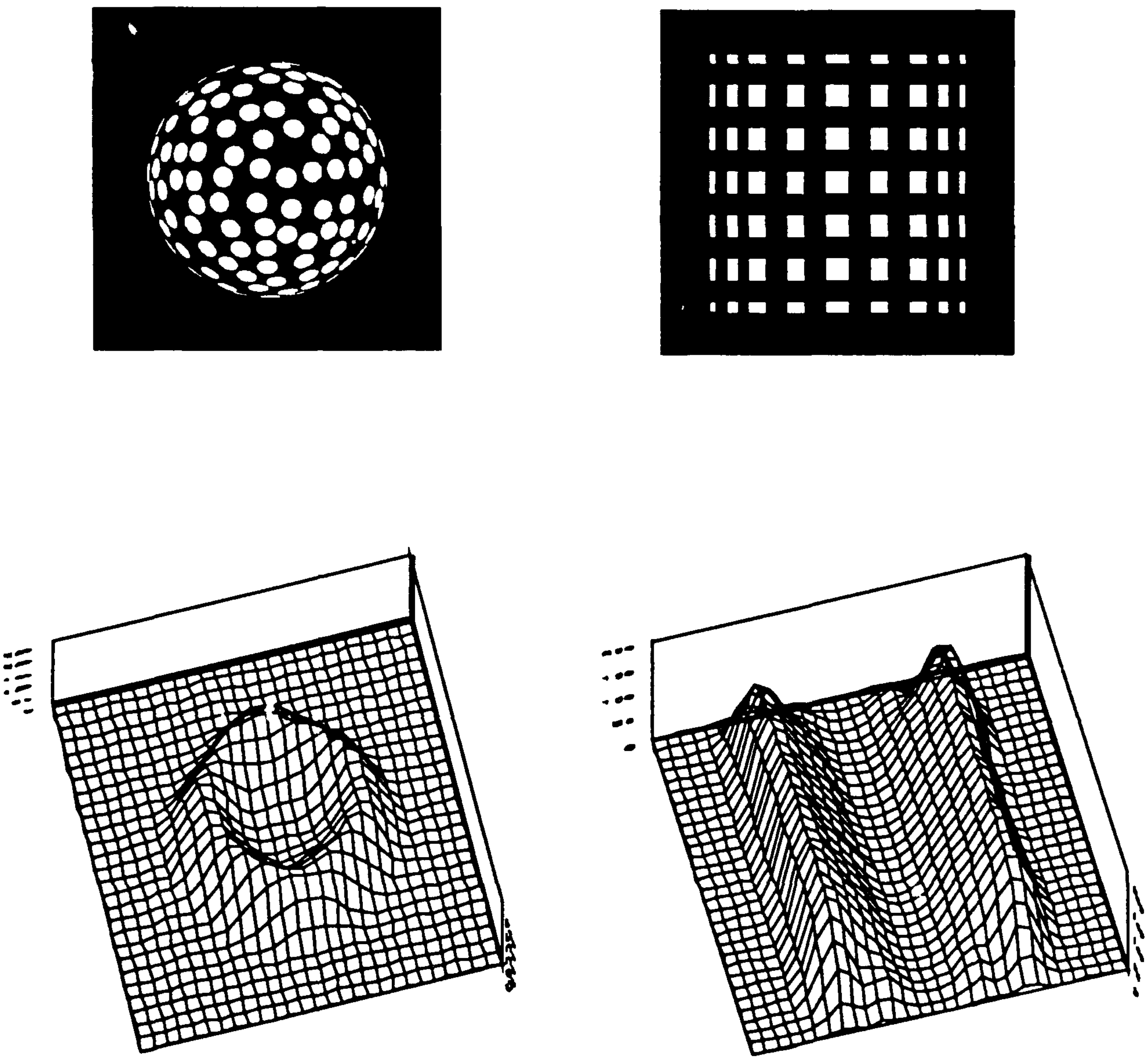

Figure 4—Shape-from-texture simulations. Upper panels show two textured stimuli which can be perceived as lying on curved surfaces (left—sphere; right—cylinder). Bottom panels show responses of network (bottom-most network in Figure 4) to these stimuli. Response plotted shows the change in surface orientation; thus, cylinder and sphere appear to curve most sharply near their edges. Network response can alternatively be viewed as signalling presence of curvature, rather than quantitative measure of amount of curvature.

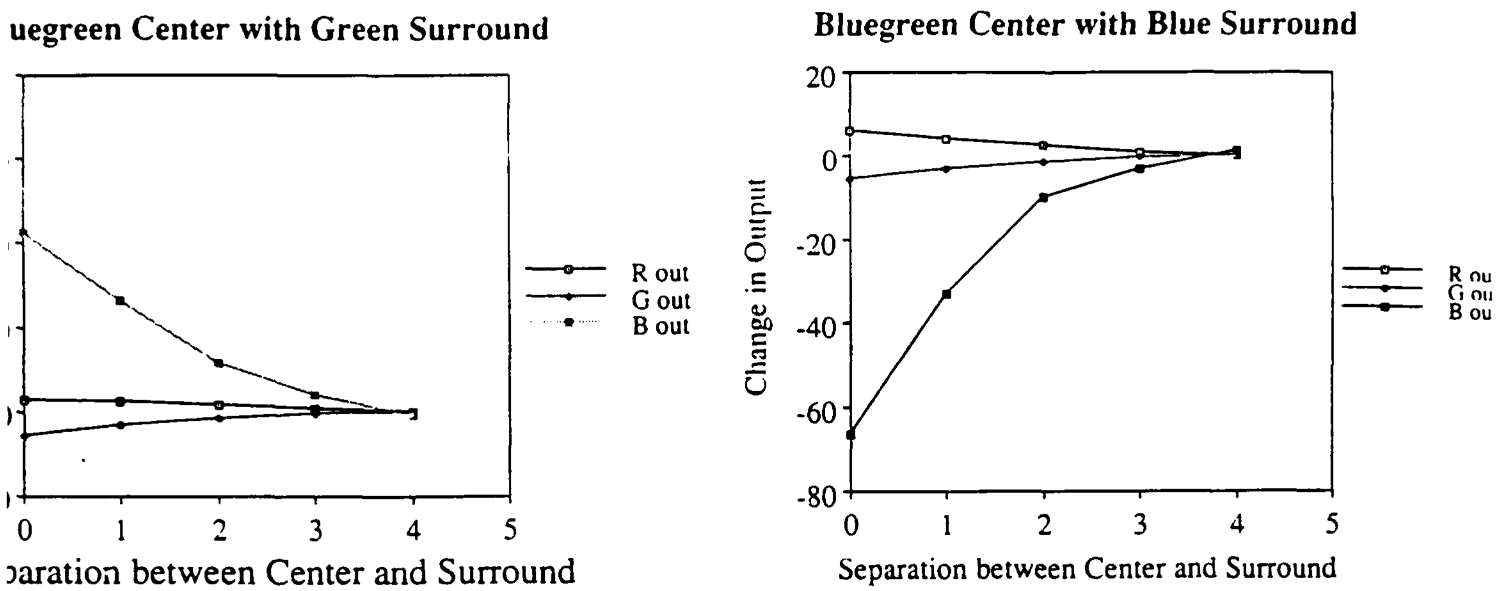

Figure 5—Color Induction. Stimuli consisting of small bluegreen spot on either a green (left) or blue (right) background were presented to color network [stimulus also had a neutral grey gap between center spot and surround]. Change in response of Red, Green, and Blue channels in network is plotted as a function of size of the gap between center and surround. Network exhibits color induction–note that green surround *increases* Blue response and decreases Green response, whereas a blue surround has the opposite effect. Simulation also shows that effect of color induction decreases monotonically as separation between center and surround increases–this conforms with psychophysical observations.

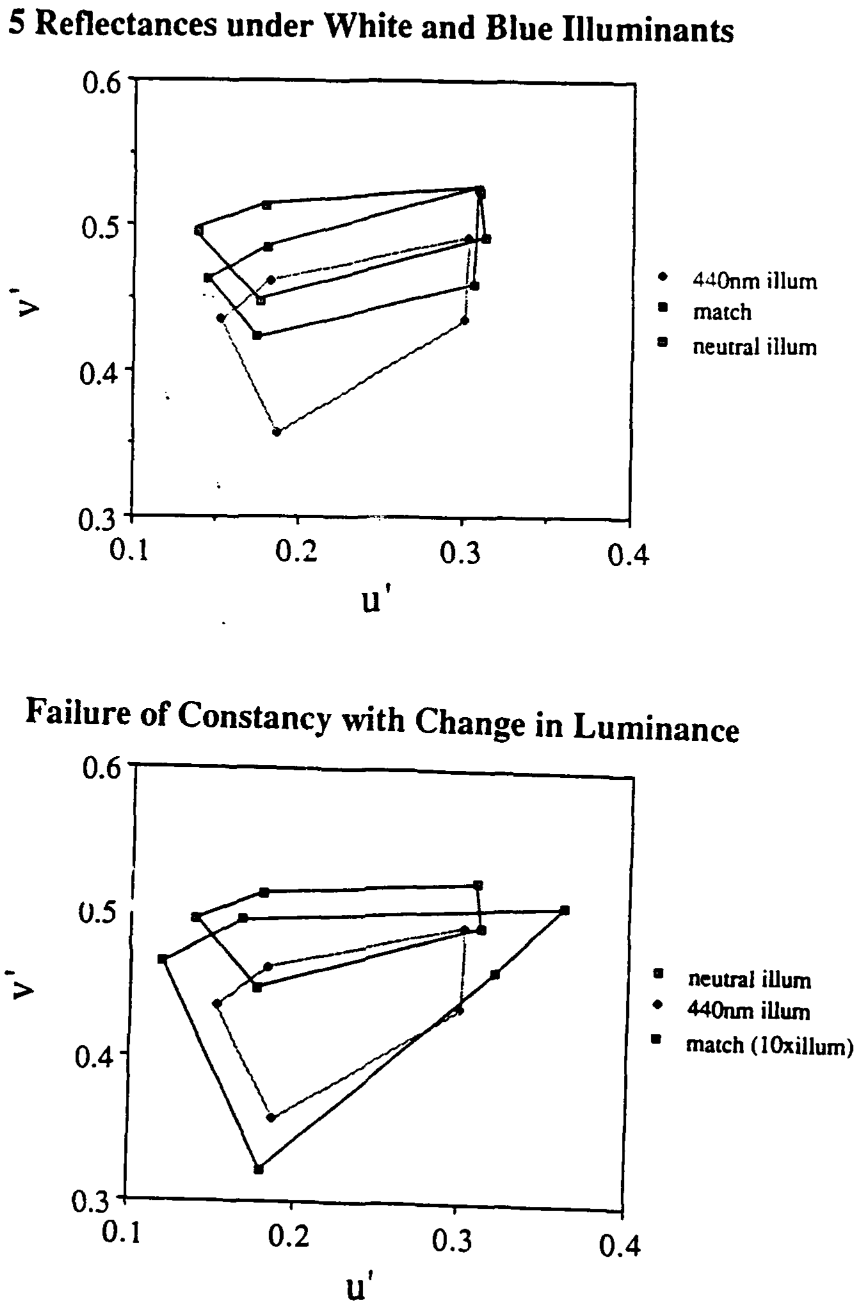

Figure 6—Role of Adaptation in Color Constancy. Simulations with color network involving 5 different colored patches viewed under different illuminants. Top plot (in u,v coordinates) shows the color constancy behavior of network–perceptual appearance of the 5 colored patches (match) is intermediate between reflectance under neutral and blue (44nm)illuminants. However, when the luminance of the illuminant is altered in addition to its hue (bottom plot) color constancy fails (note that the match reflectances are now of drastically different hue and saturation. We propose that adaptation (presumably retinal in origin) accounts for constancy under these conditions.